Ranked Set Sampling:

a combination of statistics & expert judgment

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
MAR 2012		**00-00-2012 to 00-00-2012**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Ranked Set Sampling: a combination of statistics & expert judgment	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Environmental Protection Agency ,Ariel Rios Building,1200 Pennsylvania Avenue, N.W,Washington,DC,20460**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Presented at the 9th Annual DoD Environmental Monitoring and Data Quality (EDMQ) Workshop Held 26-29 March 2012 in La Jolla, CA.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**16**	

Ranked Set Sampling

- **A sampling design where expert judgment (or simple observation) is used in combination with simple random sampling**

- **Simple random sampling is used to create a large number of potential samples. The expert then ranks these potential samples and selects which to send for analysis**

Statistician & Expert

- **Statistician:** Selects "m" sets of random samples of size "m" as potential samples (total "m x m")
- **Expert:** Within each set, the expert ranks (grades) the potential samples from highest to lowest based on the expert's opinion
- **Together:** From the first set, the largest is chosen; from the second set, the second largest is chosen; from the third set, the third largest is chosen, etc.
- **Result:** A "super-sample" of size "m"

Expert opinion can vary

- **Qualitative: Visual inspection**
  - biomass volume
  - surface soil color
  - seedling counts
  - heights of bushes
- **Quantitative: Auxiliary data**
  - historical data
  - on-site detectors
  - pH meter
  - portable equipment

How RSS works

- **Total of "m x m" potential samples will be identified and divided into "m" sets, each containing "m"**
- **One out of the "m" potential samples in each set will be sent for analysis, the rest discarded**
- **For example, if we decided to send 20 samples to the laboratory for analysis then 20 x 20 = 400 potential samples would have to be identified**

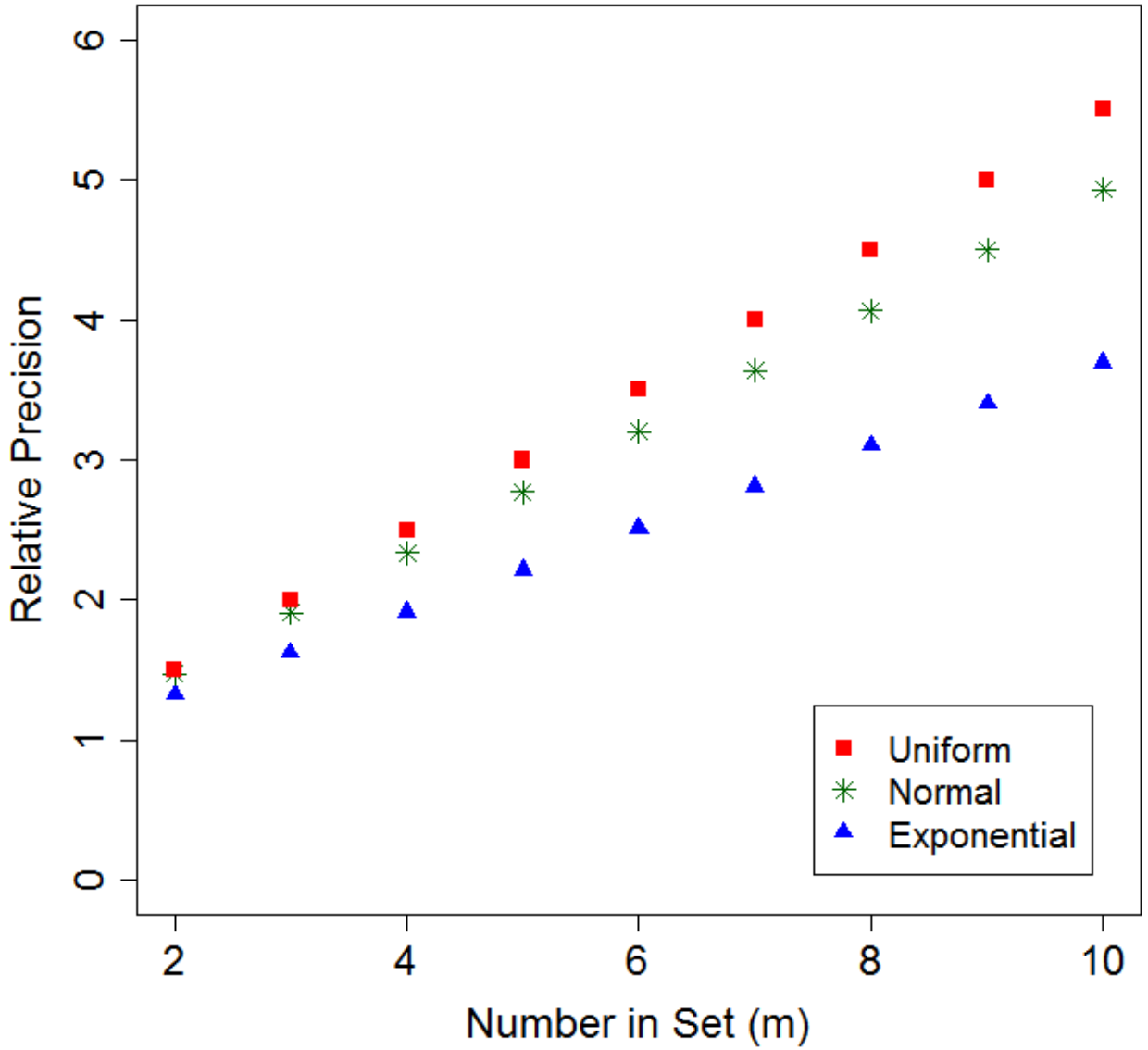
The 3-Step RSS method

- **Randomly identify "m^2" sample units using simple random sampling and allocate them into "m" sets of size "m"**

- **Rank the units within each set using the auxiliary variable selected by the expert**

- **Select the m units to be sent for analysis using this pattern:**
  - **from Set 1, select the unit with rank 1**
  - **from Set 2, select the unit with rank 2**
    …and so on until…
  - **from Set m, select the unit with rank m**

How good is this Ranked Set Sample?



RSS: The Big Advantage

- Define "advantage" as relative precision:

$$\text{Relative Precision} = \frac{\text{Variance of a random sample}}{\text{Variance of ranked set sample}}$$

  the more this exceeds 1, the better RSS

- For example: Using the preceding graph, If the distribution is Normally distributed, a RSS of 8 has the effectiveness of a simple random sample of size 8 x 4.1 = 32.8 i.e. sample of 33

- Downside is the cost of sample selection:
  - Need to identify 8 x 8 = 64 samples
  - Need to rank the 8 sets of 8 by some selected variable

Effect of Imperfect Rankings

Suppose the data are approximately Normal:

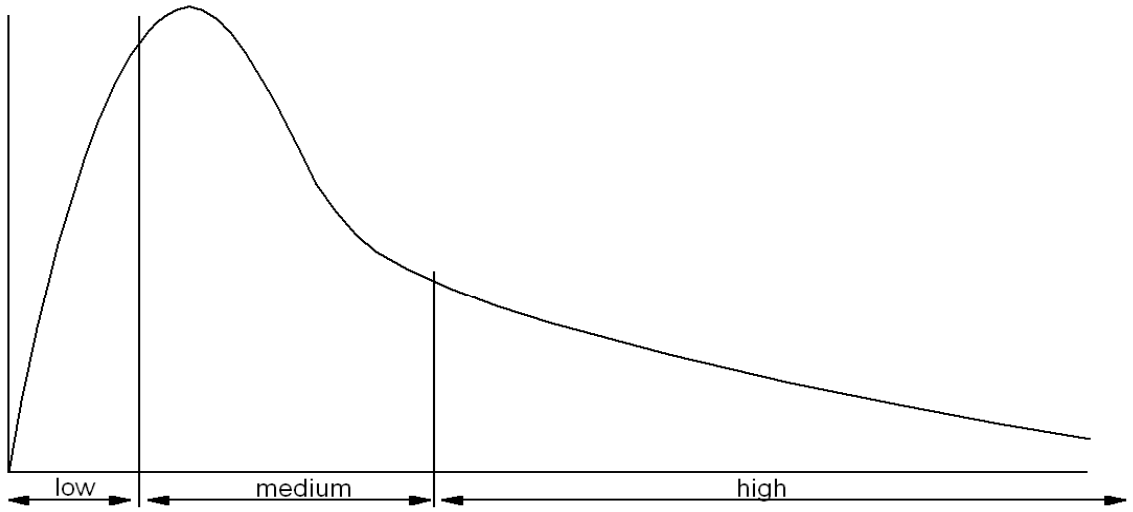
- **Small errors or confusions: Relative Precision declines up to approximately 25%**
- **Large errors or confusions: Relative Precision declines up to approximately 50%**
- **But good news! Even in the worst scenario the Ranked Set Sample is the same as a Simple Random Sample so there is no loss in statistical power**

Can't Identify m^2 samples

- Decide how many samples it is feasible to send to the laboratory (n)
- Decide how many RSS samples can be properly identified at a time (m)
- Use ordinary RSS with the "m" identified samples and repeat enough times to reach "n"
- For example, can afford to send 20 samples for analysis and can rank up to 4 units, therefore take 5 cycles of 4 RRS

Lognormal distribution of data

- Use unequal cycle allocation depending on the shape of the distribution



- More samples are taken from the higher ranked values than the lower ranked values.

Unequal Cycle Allocation: Example

For equal allocation, 12 samples needed i.e. 3 cycles of 4 sets. This distribution is skewed (lognormal) in favor of high values, thus unequal allocation is needed

Cycle Number	Set Number	Small	Medium	Large
1	1	▲	△	△
1	2	△	▲	△
1	3	△	△	▲
1	4	△	△	▲
2	1	▲	△	△
2	2	△	▲	△
2	3	△	▲	△
2	4	△	△	▲
3	1	▲	△	△
3	2	△	▲	△
3	3	△	△	▲
3	4	△	△	▲

Notice our sample of 12 consists of 3 small, 4 medium, and 5 large reflecting the known skewness of the population

Effect of unequal allocation

- **Good news: Even more effective than equal allocation (which was already much better than random sampling)**

- **But there's a cost:**
  - **Extra effort to create the "unequalness" in a meaningful way (how many extra high ones, how many low ones?)**

  - **Must adjust the way to calculate mean and variance (need to call in a statistician)**

How does it affect statistical tests?

- **If you assume the RSS is just like a random sample (which it is if RSS is ineffective) then the result:**
  - **For decision-making**
    - **~ Smaller false acceptance/rejection rates**
    - **~ Smaller "grey region" of uncertainty**
  - **For estimation**
    - **~ More accurate answers**
    - **~ Less chance of error**

Ranked Set Sampling : Conclusions

- **Pros**
  - **Better representativeness through using experts**
  - **Better precision than Random Sampling**
  - **Same simple formulae to use**

- **Cons**
  - **Increased cost of the expert ranking samples**
  - **Difficulty quantifying exact improvement**
  - **Need to find best variable to do the ranking on**

…but the Pros definitely outweigh the Cons!

Further advice on Ranked Set Sampling

- ***Guidance on Choosing a Sampling Design for Environmental Data Collection QA/G-5S*** **(www.epa.gov/quality)**